

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

E

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平6-75996

(43) 公開日 平成6年(1994)3月18日

(51) Int.Cl.<sup>5</sup>

G 0 6 F 15/40

識別記号

5 1 0 J

庁内整理番号

7218-5L

F I

技術表示箇所

審査請求 未請求 請求項の数1(全 4 頁)

(21) 出願番号 特願平4-229587

(22) 出願日 平成4年(1992)8月28日

(71) 出願人 000001889

三洋電機株式会社

大阪府守口市京阪本通2丁目5番5号

(72) 発明者 茂木 健

大阪府守口市京阪本通2丁目18番地 三洋  
電機株式会社内

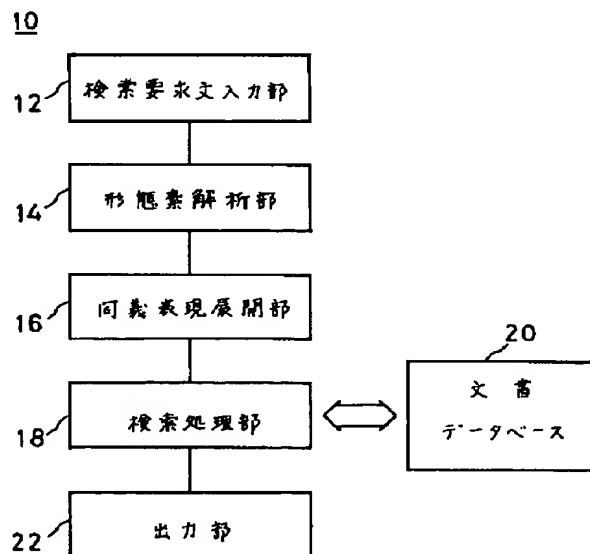
(74) 代理人 弁理士 山田 義人

(54) 【発明の名称】 文書検索装置

(57) 【要約】

【構成】 検索要求文入力部12に入力した検索要求文について、形態素解析処理部14で形態素解析処理を施して、同義表現展開部16に与える。同義表現展開部16では、活用変化、複合語の分解、品詞変化および文体変化に対応できるような同義表現を生成して、検索処理部18に与える。検索処理部18では、検索要求語およびその同義表現について文書データベース20と照合し、照合に成功すれば対応する文章を文書データベース20から引き出して、出力部22から出力する。

【効果】 自由度が高くかつユーザの検索意図に沿った文章を検索できる。



1

## 【特許請求の範囲】

【請求項1】入力される検索要求語について形態素解析を施す形態素解析手段、

前記形態素解析した前記検索要求語の同義表現を生成する同義表現展開手段、

文章が格納される文書データベース、

前記検索要求語および前記検索要求語の同義表現に基づいて前記文書データベースから対応する文章を検索する検索処理手段、および前記文書データベースからの対応する文章を出力する出力手段を備える、文書検索装置。

## 【発明の詳細な説明】

【0001】

【産業上の利用分野】この発明は文書検索装置に関し、特にたとえば検索要求語と類似する表現の文章を検索するために用いられる、文書検索装置に関する。

【0002】

【従来の技術】従来の文書検索装置では、キーワードを入力し、文書データベースを参照して検索を行い、キーワードの完全一致を条件として、対応する文書が引き出されていた。

【0003】

【発明が解決しようとする課題】しかし、この従来技術では、文書データベース側にキーワードそのものが入っていないと対応する文書を引き出すことができなかった。そこで、より広範囲に類似の文書を引き出すために、特開平3-130873号公報の技術が提案されている。

【0004】この従来技術では、見出しの漢字と同義な意味を有する漢字および類義な意味を有する漢字を記憶した類義字辞書を有し、入力文のキーワードをその類義字にまで展開して検索する。しかし、この技術では、検索要求語と文章表現自体が類似する文章を引き出すことはできなかった。それゆえに、この発明の主たる目的は、検索要求語と類似する文章を引き出すことができる、文書検索装置を提供することである。

【0005】

【課題を解決するための手段】この発明は、入力される検索要求語について形態素解析を施す形態素解析手段、形態素解析した検索要求語の同義表現を生成する同義表現展開手段、文章が格納される文書データベース、検索要求語および検索要求語の同義表現に基づいて文書データベースから対応する文章を検索する検索処理手段、および文書データベースからの対応する文章を出力する出力手段を備える、文書検索装置である。

【0006】

【作用】入力された検索要求語について形態素解析手段で形態素解析を施し、同義表現展開手段に与える。同義表現展開手段では、たとえば活用形への展開、複合語の分解、品詞変化あるいは文体変化などを考慮して、検索要求語の同義表現を生成する。そして、検索処理部で、

2

検索要求語およびその同義表現と文書データベースとを照合して文章検索を施す。そして照合に成功すれば、検索要求語に対応する文章を文書データベースから引き出して出力する。

【0007】

【発明の効果】この発明によれば、検索要求語の言語的類似性を考慮して、検索要求語と類似する文章を文書データベースから引き出すことができるので、自由度の高い文章検索が可能となる。また、たとえば対訳表現集の検索等、少量文書のきめ細かい検索に有効となり、さらに、ユーザの検索意図に沿った文章を検索できる。

【0008】この発明の上述の目的、その他の目的、特徴および利点は、図面を参照して行う以下の実施例の詳細な説明から一層明らかとなろう。

【0009】

【実施例】図1を参照して、この実施例の文書検索装置10は、検索要求語の言語的類似性を考慮して文章を検索するものである。文書検索装置10は、検索要求語を入力する検索要求語入力部12を含み、入力された検索要求語は形態素解析部14で形態素解析を施された後、同義表現展開部16に与えられる。同義表現展開部16では、語句の分裂、語句の結合、たとえば名詞形と動詞形との相互間の品詞の違い、時制の違いおよび活用変化などに対応できるように、検索要求語の同義表現を生成する。そして、検索処理部18で、検索要求語および同義表現について、文章が格納された文書データベース20を参照して文章検索を行う。照合に成功すれば、検索要求語に対応する文章を文書データベース20から引き出して、出力部22から出力する。

【0010】ここで、具体例を挙げて説明する。入力された検索要求語に対し、以下の操作1～4によってその同義表現を生成し検索する。これにより、語句の分裂、語句の結合、品詞の違い、時制の違い、常体／敬体の違い、平叙形／疑問形の違いおよび活用変化などを吸収した検索が可能になる。

# 1. 複合語の分解、句の分解（語句の分裂への対応）

a) 検索要求語K<sub>1</sub>が、A、Bの結合した複合名詞、「AのB」または「AをBする」の形の句であるとき、同義表現K<sub>2</sub>をA、BのANDとして検索する。

【0011】

K<sub>1</sub> = 「AB」、「AのB」、「AをBする」

K<sub>2</sub> = 「A」 \* 「B」

（ここで、ABはA、Bの結合した複合名詞、A \* BはA、BによるAND検索を行うための同義表現を表す）

例1) 検索要求語K<sub>1</sub> = 「ディレクトリ表示」

同義表現K<sub>2</sub> = 「ディレクトリ」 \* 「表示」

照合成功文 = 「ディレクトリ構造の表示」

例2) 検索要求語K<sub>1</sub> = 「文書の検索」

同義表現K<sub>2</sub> = 「文書」 \* 「検索」

照合成功文 = 「文書」をソフトウェアにより検索す

3

る。」

b) 検索要求語  $K_1$  が、A、Bの結合した複合名詞、「AのB」または「AをBする」の形の句であるとき、A、Bのそれぞれを同義表現として検索する。

【0012】

$K_1$  = 「AB」、「AのB」、「AをBする」

$K_{21}$  = 「A」、 $K_{22}$  = 「B」

例) 検索要求語  $K_1$  = 「構造の分析」

同義表現  $K_{21}$  = 「構造」、 $K_{22}$  = 「分析」

照合成功文 1 = 「文の意味構造が表されている。」

照合成功文 2 = 「動作の主体であると分析できる。」

## 2. 複合名詞化 (語句の結合への対応)

検索要求語  $K_1$  が、「AのB」「AをBする」の形の句であるとき、同義表現  $K_2$  をA、Bの結合した複合名詞とし検索する。

【0013】  $K_1$  = 「AのB」または「AをBする」

$K_2$  = 「AB」

(ここで、A、Bは名詞、ABはA、Bの結合した複合名詞を表す)

例) 検索要求語  $K_1$  = 「文書の検索」

同義表現  $K_2$  = 「文書検索」

照合成功文 = 「文書検索における処理」

## 3. 助動詞、助詞の削除 (品詞の違い、時制の違い、常体/敬体の違い、平叙形/疑問形の違いへの対応)

検索要求語  $K_1$  が、サ変動詞「Aする」または「A+助動詞、助詞(ている/させる/られる/た/です/ます/か、など)」で終了する語句であるとき、助動詞を削除したものを同義表現  $K_2$  とし検索する。

【0014】  $K_1$  = 「XAする」、「XAさせる」、「XAです」など

$K_2$  = 「XA」

(ここで、Xは任意の語句を表す)

例1) 検索要求語  $K_1$  = 「処理する」

同義表現  $K_2$  = 「処理」

照合成功文 = 「自然言語処理」

4

例2) 検索要求語  $K_1$  = 「これが問題です」

同義表現  $K_2$  = 「これが問題」

照合成功文 = 「要するに、これが問題である。」

例3) 検索要求語  $K_1$  = 「時間がある」

同義表現  $K_2$  = 「時間が」

照合成功文 = 「これを全部やる時間がありますか。」

## 4. 活用語尾削除 (活用変化への対応)

10 検索要求語  $K_1$  が、用言(動詞/形容詞/形容動詞)で終了する語句であるとき、その用言から活用語尾を削除したものを同義表現  $K_2$  と設定し検索する。

【0015】  $K_1$  = 「XA」

$K_2$  = 「XA'」

(ここで、Xは任意の語句、Aは用言、A'はその語幹を表す)

例) 検索要求語  $K_1$  = 「努力を怠る」

同義表現  $K_2$  = 「努力を怠」

照合成功文 = 「努力を怠らなければ、必ず成功する。」

20 そして、照合成功文が文書データベース20から引き出され、出力部22から出力される。このようにして、検索要求語および上述の操作1~4のいずれかによって生成された同義表現に基づいて照合成功文を検索できる。

【0016】なお、上述の例の中には、照合成功文が単語や句の状態のものが存在するが、その単語や句を含む文章が文書データベース20から引き出されることに留意されたい。

【図面の簡単な説明】

【図1】この発明の一実施例を示すブロック図である。

30 【符号の説明】

10 …文書検索装置

14 …形態素解析部

16 …同義表現展開部

18 …検索処理部

20 …文書データベース

【図1】

